When using the structured data model, the data is subject to previously known limitations on the type and length of each attribute; the data structure is known and defined using the data schema; its automatic modification during the model operation is difficult. Interpreting data without knowing the scheme is impossible. An example of a structured data model implementation can be any relational database management system (DBMS).

Developing a model for unstructured data is extremely difficult for the following reasons: data is usually presented in a natural language, which makes it difficult to work with it; the complete absence of a certain structure imposes serious restrictions on possible operations with data. The automatic allocation of the structure in such data, as a rule, cannot be performed in an unambiguous way.

Semi-structured data is any intermediate data between structured and unstructured. The structure in such data may be incomplete, underdetermined, and also permit exceptions. When working with data, the degree of its correctness is not known in advance, and, as a result, the necessary tools are also unknown for assessing the 'correctness' of the data. Thus, there must be an exception handling tool in this model that allows you to formulate a method for querying this data based on predefined criteria.

The paper analyzes the three main methods for presenting semi-structured information:

1) OEM (Object Exchange Model);
2) XML (Extensible Markup Language);
3) RDF (Resource Description Framework)

The XML language is a subset of the SGML language (Standard Generalized Markup Language). SGML is a system for describing structured document types and markup languages for document copies of such types. This language allows you to divide any document into two logically independent parts; one of them defines the structure of the document, and the other contains the text itself. The structure definition is called the Document Type Definition (DTD).

MATHEMATICAL METHODS OF BIG DATA REPRESENTATION

Kateryna Bielienkova, student
Maryna Bulaienko, Associate Professor, PhD in Technical Sciences, Research Advisor
Viktoriia Buhaieva, Senior Teacher, Language Consultant
O. M. Beketov National University of Urban Economy in Kharkiv

K. Lynch, D. Laney define Big Data as a set of methods and tools for processing structured and unstructured heterogeneous dynamic data of large volumes with the purpose of their analysis and use to support decision making.

Examples of sources of origins of big data can be continuously incoming data from measuring devices, meteorological data, data of remote sensing of the

Earth, data streams on the location of subscribers of cellular networks, devices for audio and video recording, etc. The beginning of the widespread use of these sources has led to the penetration of big data technologies into research and development, into the commercial sector and into public administration.

Big Data is an alternative to traditional database management systems and business intelligence class solutions. This class also includes parallel data processing tools (NoSQL, MapReduce, Hadoop algorithms). According to the DCA (Data-Centric Alliance) company, Big Data is understood not as a specific amount of data and even data, but processing methods that allow to distributedly process data. These methods can be applied both to large data arrays (such as data from all pages on the Internet) and to small data arrays (information about the daily goods received by the store).

The aim of the paper is a formal description of various mathematical models of data representation, such as a multidimensional model, an object model and a graph model. Selection of operations and carriers, as well as the ways of using them.

The basic concepts of the multidimensional data model are:

- hypercube of data (G),
- measurement (V),
- attribute (A),
- cell (X),
- value of (G (V, A)).

A hypercube of data contains one or more dimensions and is an ordered set of cells. Each cell is defined by one set of dimension values which are attributes. A cell can contain data or be empty.

Measurement is understood as a great number of attributes that form one of the faces of the hypercube. An example of a time dimension is a list of days, months, quarters. An example of a geographical dimension can be a list of territorial objects: localities, regions, districts, regions, countries, etc.

Thus, a hypercube of data can be designated as a set of cells that corresponds to the sets V, A: G (V, A). Hypercubes maintain a hierarchy of dimensions and formulas without duplicating their definitions. A set of corresponding hypercubes constitutes a multidimensional database (or data storage). It is good to use multidimensional data representation for data visualization tasks and their analysis, but due to the sparseness of the hypercube, the amount of data in that case is larger compared to the relational representation.